A Framework for Analyzing AR/VR Collaborations

An Initial Result[1]

Michael Tanaya, KouKeng Yang, Taran Christensen, Shen Li, Michael O'Keefe, James Fridley, Kelvin Sung
Computing and Software Systems
University of Washington Bothell
Bothell, USA

***Abstract*—The recent advances in Virtual Reality (VR) and Augmented Reality (AR) research have enabled investigations into collaborative efforts that equip participants with heterogeneous technological configurations in various degrees of immersion. Results from this work are interesting and may seem futuristic which can cause challenges in understanding the underlying impacts and gaining insights. This paper describes an attempt in identifying the fundamental elements of a digital interaction and remote collaboration; and proposes the *Cross Reality Collaboration Framework* (CRCF) to provide a unified platform for discussing, comparing, and contrasting these efforts. Though in an early stage of development, the current iteration of CRCF is capable of analyzing configurations that are based on drastically different technologies (e.g., an AR application vs. a location aware mobile application), comparing them, and revealing insights. This paper demonstrates the potentials of CRCF by applying it to analyze popular technological configurations and identifying opportunities for investigations. The paper then describes a prototype solution addressing the identified opportunity and presents results from a testing environment involving four collaborators with distinct technological configurations.**

Keywords—Virtual Reality; Virtual Space; Augmented Reality; Mixed Reality; Collaboration

I. INTRODUCTION

Virtual Reality (VR) and Augmented Reality (AR) research has gone through resurgent cycles over past decades [1], [2]. The most recent commercial product releases, e.g., the HTC Vive[2], or the Microsoft HoloLens,[3] are indications that the technologies are maturing and bringing their potentials for significant impacts to the general public [3].

As the field continues to develop, VR/AR technologies are increasingly being relied on as foundations for investigations *across reality spaces* where solutions span the continuum between *virtual* and *physical* spaces [4]. One of the critical areas of study is in the understanding of the necessary support and implications of remote collaborations where participants may be equipped with heterogeneous technological configurations. For example, connecting distant VR participants with others who have access to wall-size displays for visualizing data [5], or providing remote experts a near-immersive interaction with explorers in the fields [6].

This paper proposes the *Cross Reality Collaboration Framework* (CRCF) as a platform for analyzing and discussing work done involving digital interactions over remote distances that span virtual and physical spaces. The CRCF analyzes a technological configuration according to the reality spaces of its user; objects involved; input, camera, and display mechanisms; and interaction space. Though in an early stage of development, the current version of CRCF can serve as a unified platform for analyzing single-machine and web-based applications, VR/AR equipped systems, and location-aware mobile apps. Such analysis provides insights into shortcomings, potential improvements, and opportunities for further investigations.

Based on a CRCF analysis of some currently popular configurations, opportunities are identified, and a network infrastructure is developed to support future studies of remote collaboration with on-site virtual presence. This paper presents the results from an initial testing environment that supports collaboration among participants equipped with VR, AR, distance sensors, and traditional primary computing environments.

This paper first presents a background context for the CRCF development. The section following that introduces the framework with a discussion on the current limitations and an example analysis illustrating its potentials. Section four describes a prototype implementation based on the analysis results, where section five presents the results. The paper concludes with a summary of potentials for future work.

II. BACKGROUND

Sutherland reported their head-mounted 3D system, furnished with stereoscopic displays and head positional tracking capabilities, in 1968 [7]. To this system, Clark integrated a controller with six degrees of freedom and created an elementary immersive Virtual Space that supported the designing of 3D surfaces [8]. Commercial uses of VR dated as far back as 1980's with the arcade video game Battlezone[4] [9]. Closely tied to the development of the supporting technologies, the research in VR underwent periods of renewal in popularity

---

[1] This work was supported in part by generous grants from Microsoft External Research under the Computer Gaming Curriculum in Computer Science RFP, Award Numbers 15871 and 16531.

[2] https://www.vive.com/

[3] https://www.microsoft.com/microsoft-hololens/

[4] https://en.wikipedia.org/wiki/Battlezone_(1980_video_game)

[1], [10]. Many believe the most recent technological advancements are going to bring VR to the general public with a substantial impact [3].

With its transparent CRT display, Sutherland's head-mounted system displayed computer generated virtual objects in front of the surrounding physical world [7]. In this way, this system can also be considered as the first AR device. Similar to the case of VR, AR investigations have also gone through similar resurgence [2], [11], [12]. Most recently, the HoloLens [13] has demonstrated the maturity of markerless AR [14] as a commercial product and has enabled much research and many interesting results [4].

As the fields mature, VR/AR systems are becoming the foundations for next generation investigations where diverse technologies are integrated as part of the solutions. For example, Chen et al. integrated markerless AR with video conference capabilities and brought remote experts into distant or potentially difficult to access locations for on-site analysis [6]; the DIGISCOPE project provided the infrastructure to support remote collaborators working in immersive VR and wall-size display systems [5]; and Sra and Schmandt created a multiuser system that allowed the physical world to be used as a template for the construction of VR spaces [15].

These are examples of collaborations *across realities* where the solution involves heterogeneous combinations of VR/AR elements, overlaps and combines virtual and physical spaces, and connects participants over geographic distances. In the increasingly connected and digitized world, this area of investigation will continue to grow and flourish. A framework for analyzing, comparing, and contrasting these solutions is essential for understanding the relationships between the on-going work and identifying new directions for investigation.

III. THE CROSS REALITY COLLABORATION FRAMEWORK

The *Cross Reality Collaboration Framework* (CRCF) is designed to describe the technological configuration and the interaction space participants in a collaboration. Although part of the framework is still under study and refinements, an analysis of current popular technologies based on the relatively stable portion of the framework reveals insightful observations and has led to the investigations and results presented in this paper.

In the context of CRCF, elements of a configuration are classified as *virtual* or *physical* according to if they can be felt or sensed with or without the assistance of technologies. When an element is *virtual*, it can be considered as *logical* and existed only in the digital space where the specifics are enabled and programmed by the supporting technology; while a *physical* element is *real matter* and interpreted accordingly by the real-world physics. For example, the document edited by a word processor is *virtual*, while the printed version is *physical*; or the point of view in a VR system is programmable and is thus *virtual*, while the point of view of an AR system exists in the physical world and is thus *physical*.

A. Elements of a Digital Interaction

The following focuses on participants who interact through the assistance of digital technology in accomplishing tasks. In such interactions, the following elements can be identified.

- The **User** (U): the participant; e.g., the editor of a word processor is the user.
- The **Objects** (O): the tangible entities; e.g., the document being edited is the object.
- The **Control** (C): the mechanics upon which the user manipulates the objects; e.g., the keyboard and mouse used while editing a document.
- The **Point of View** (P): a portion of the solution space; e.g., the window into the working document.
- The **Display** (D): a presentation of the point of view; e.g., computer monitor presenting the window.[5]

The *Cross Reality Collaboration Framework* (CRCF) classifies the above elements as *virtual* or *physical* according to the availability of assistance from technologies and their operating spaces. For example, a person editing a word processing document can be described as—physical user (U_p), virtual object (digital file, O_v), physical control (typing on keyboards, C_p), virtual point of view (a window, P_v), and physical display (computer monitor, D_p); or, $\{U_p, O_v, C_p, P_v, D_p\}$. In general, if supported, automatic spelling correction can be considered as a virtual user (bot). This means that the interaction involves both virtual and physical users and thus, the user is U_{v+p}. As another example, in its current and simplest form, a VR application interaction involves a physical user (U_p) manipulating computer graphics objects (O_v) with physical wands (C_p) where the world is viewed from a programmable position (P_v) and shown on the physical head-mounted display (D_p). In this way, the interaction can be described as $\{U_p, O_v, C_p, P_v, D_p\}$.

It is interesting, though not surprising, that according to the CRCF, a word processor and a VR application can potentially be identical—the VR application can very well be a word processor. The missing element is the degree of **Immersion**, I, that the user experienced. In general, it is challenging to articulate whether an immersive experience is virtual or physical. For this version, the I element will only be approximated subjectively as yes or no. As will be pointed out, this element can be leveraged for identifying areas of investigation.

The relationship between the control (C) element and the existing user interface paradigms is currently under investigations. It is essential to articulate a virtual or physical classification in the context of existing Natural User Interface [16] and Tangible User Interface [17] paradigms. The current

---

[5] For convenience, the discussion focuses only on one of the human senses, the sight. In general, P (point of view) and D (display) can be replaced by terms that describe the sampling and reproducing of information for other forms of human senses, e.g., hearing.

Example	Elements of Interaction			S	I[a]	N[b]
	U	O	P			
Microsoft WORD	v+p	v	v	v	n	n
Google Doc	v+p	v	v	v	n	y
Google Cardboard	p	v	v	v	n	n
HTC Vive	p	v	v	v+p	y	n
Hololens	v+p	v+p	p	p	y	n
Pokémon Go	p	v	p	p	n	y

[a] Support for Immersive Experience. [b] Support for multiple participants

Table 1: Applying the stable portion of the CRCF

technology restricts the display (D) element to be always physical, e.g., a hologram may be an example of virtual displays. For these reasons, the C and D elements will be omitted from the discussion for the rest of this paper.

B. The Collaborative Space

The definition of O (objects) demands that tangible entities be involved in the digital interaction. This requirement excludes certain categories of collaborative efforts; e.g., a conversation about an idea that does not produce tangible results would be excluded. However, it is also true that with appropriate solutions for representing, storing, and sharing, the existence of tangible objects can potentially liberate both the time and space constraints in a collaboration [18]. Participants can analyze and modify shared tangible objects at anytime from anywhere.

The **Space** element, S, is included to support the discussion of collaborative efforts that span the virtual and physical continuum [6], [15].

C. Understanding The Stable Portion of the CRCF

Applying the CRCF, Table 1 summarizes the results of analyzing the elements of interactions (the columns) while working with popular applications or technologies (the rows). Notice the two columns to the left that summarize the support for immersive experiences (I) and multiple users (N).

The U_{v+p} of Google Doc and HoloLens refer to the auto spell-checker and Microsoft's Cortana. The O_{v+p} on the HoloLens reflects that this is the only configuration that supports its users interacting with both virtual and physical objects in the environment. The S_{v+p} of the Vive row recognizes the fact that with the distance sensors and location mapping, a user's physical movements in the real world can become an integral part of the solution space in the virtual environment.

D. Observations

There are multiple approaches to analyze Table 1, and each can lead to additional investigations. For example, considering the potentials of an element to be both virtual and physical. A U_{v+p} can be understood as the support of intelligent and autonomous decisions (AI) provided by the respective technology. The potential meanings of O_{v+p} and S_{v+p} have been discussed previously. A P_{v+p} for configurations with a physical point of view, e.g., the HoloLens, can imply a virtual minimap-like point of view. However, it can be challenging and potentially interesting to articulate the meaning and purpose of integrating a physical point of view for a Vive user.

Another approach to analyzing Table 1 would be examining the options, potentials, or possibilities of exchanging the virtual and physical classifications of an element. For example, replacing the S_p of a HoloLens, or an AR environment, with a S_v would mean allowing a remote user to explore and interact with physical and virtual objects in an environment by manipulating a physical point of view; something that can be achieved with a remote-operated robot. Currently, the authors are exploring alternative solutions based on mapping the physical space.

Insights can also be gained when comparing the rows and columns of Table 1. For example, the last two rows indicate that the difference between a simple application developed for the HoloLens differs from Pokémon Go[6] in three aspects: the relationship between virtual and physical objects (actual interaction vs. simple overlap), the immersive experience, and, the support for collaboration. While the non-interaction of the virtual and physical objects is a severe limitation, this observation does suggest investigations for supporting an AR-like experience by trading-off the immersion factor for cost.

Examining the combinations of the O (objects), P (point of view), and S (space) columns suggest that independent of immersion, VR and AR can be considered as studies of approaches to integrate *virtual* and *physical* realities in these elements. Once again, the immersion factor can be articulated as an optimization for cost, serving as an inspiration for the investigation into providing physical and virtual interactions at a lower cost.

IV. PROTOTYPE INVESTIGATION

Inspired by the potentials of trading-off cost for immersive experience, in parallel with the observation that the proper representation, storage, and sharing of tangible objects (O) can alleviate the time-space constraints in collaborations, the *Cross Reality Collaboration Sandbox* research group at the University of Washington Bothell has begun the investigation into prototyping a solution for representing the physical environment, and developing an infrastructure for storing and sharing objects to support distant collaborations. The modest initial goals include verifying the CRCF ability in describing heterogeneous interaction elements; and exploring approaches to reproduce CRCF signatures with more economical configurations. Ultimately, the team's goal is to investigate and facilitate collaborations between participants that are located in a physical space of interests with remote participants. For example, a collaboration to virtually and physically decorate a house where the homeowners are located in the house and the interior designers are located off-site.

A. Technology and Representation of Physical World

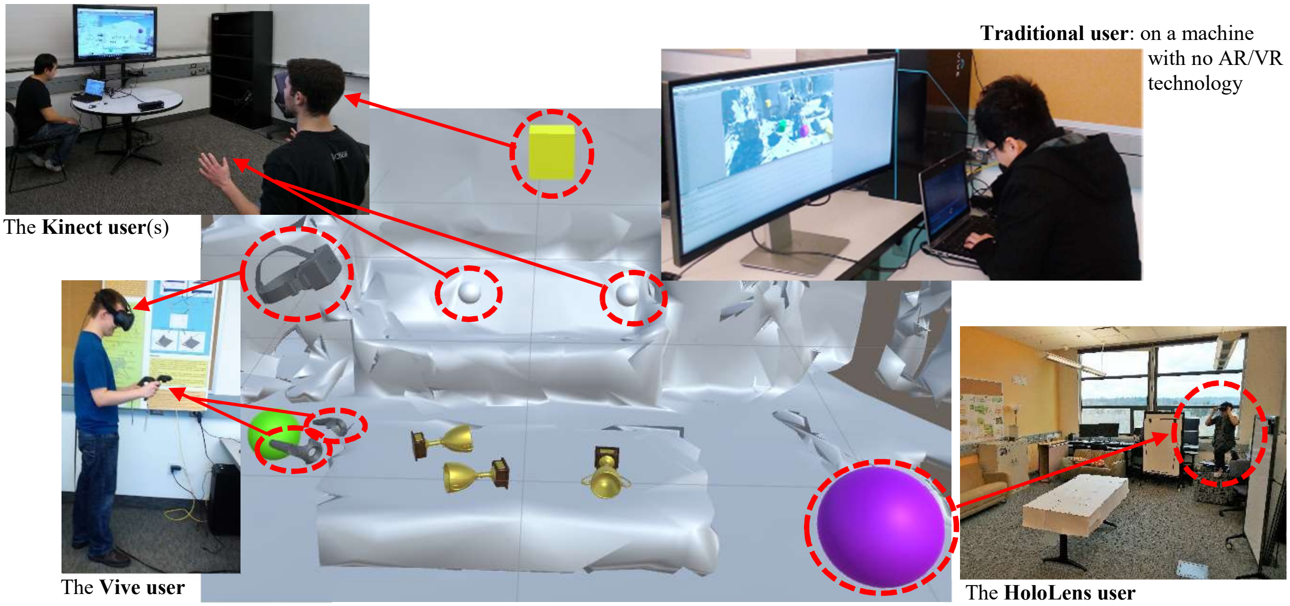
The Unity3D[7] platform was chosen as the foundation for implementation because it supports a wide variety of existing VR/AR systems. The HoloToolkit[8] Spatial Mapping library was

---

[6] http://www.pokemongo.com/

[7] https://unity3d.com/

[8] https://github.com/Microsoft/HoloToolkit-Unity


Traditional user: on a machine with no AR/VR technology

The **Kinect user**(s)

The **Vive user**

The **HoloLens user**

Screen capture from the **Traditional user's** display

Fig 1. The Collaboration Environment.

adopted to provide a trivial approximation of the physical environment through simple meshes. In this way, the physical and virtual objects in the collaboration environment are both represented by traditional 3D geometries. The collaborators that are on-site can choose to switch off the display of spatially mapped meshes.

B. *The Augmented Space Library*

An efficient networking library is a straightforward solution for *storing* and *sharing* objects, and facilitating collaborations across geographic distances. Early investigations into shareable VR spaces began as early as the late 1980s [19]. More recent approaches considered separately on representing Virtual Space based on existing geometries [20], and communicating mapped physical world [21]. The *Augmented Space Library* (ASL) combines these ideas and supports the augmentation of existing 3D geometries onto communicated scanned world meshes. A custom server is developed to support the communication of scanned meshes to all collaboration participants, while the maintenance and synchronization of defined geometries are based on the Photon Unity Networking package.[9]

C. *The Participant Applications*

Based on the ASL, four separate simple applications are developed. First two applications are for the existing popular VR (e.g., Vive) and AR (e.g., HoloLens) technologies. A third application attempts to reproduce the Vive CRCF signature of $\{O_v, P_v, S_{v+p}\}$ with the cheaper but non-immersive Microsoft Kinect sensor. The fourth and last is a technology independent application designed to demonstrate the viability of collaborating in this environment as a *traditional* user.

V. RESULTS

Figure 1 depicts the environment designed for verifying the functionality of the ASL and for demonstrating the versatility and convenience of CRCF in explaining collaborators with diverse technological configurations. This testing environment is constructed by a quick scan of our lab with the HoloToolkit Spatial Mapping system, followed by distributing the scanned meshes to all collaborators. The center image of Figure 1 highlights three of the collaborators, beginning from the left: Vive (left), Kinect sensor (top-left), and HoloLens (bottom-right). The fourth collaborator participates via a *traditional* configuration without any special technologies. The center image is the screen capture from the display of the *traditional user*. Note that this participant is not immersed in the collaboration environment and we have chosen to not show his presence.

A. *The ASl LibraryFunctionality*

Figure 2 depicts the view from the HoloLens user that corresponds to the center image of Figure 1. The dotted circles highlight the locations of the Vive user headset and control,

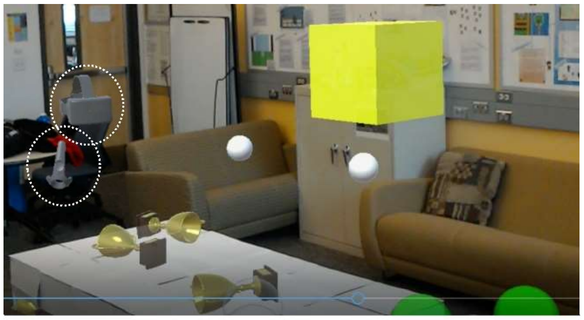


Fig. 2. View from the HoloLens User

---

[9] http://doc.photonengine.com/en/pun/current/tutorials/photon-unity-and-networking-links

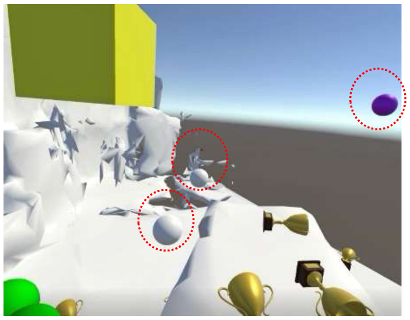


Fig. 3. View from the Vive User

while the yellow cube and the two gray spheres are the head and hands positions of the Kinect user.

Figure 3 is a screen capture of the Vive user of the same test environment. From this view, the top-left yellow cube and the two highlighted gray spheres represent the head and hands locations of the Kinect user, while the highlighted top-right purple sphere is the position of the HoloLens user. Notice that because of the relatively narrow fields of view of the HoloLens device, as depicted in the bottom-right photo of Figure 1 and the position of the purple sphere in Figure 3, the HoloLens user has to be relatively far away from the table for the device to capture the image in Figure 2.

The bottom photograph of Figure 4 shows the configuration of the Kinect user: standing in front of the Kinect sensor, viewing the environment from the TV, and interacting with the virtual collaboration environment via real-world physical movements. For example, the user can reach a hand forward in the physical space to grab a trophy in the virtual environment, or walk backward in the physical space to move away from the table in the virtual environment. The top image in Figure 4 is a screen capture of the same environment. In this case, in addition to the highlighted Vive (left) and HoloLens (right) users, the Kinect user himself (center) is also shown in the display as the highlighted dark gray skeleton and hands being shown as green spheres. The Kinect user refers to this skeleton when performing physical movements to interact with the environment. Notice that it is possible to include multiple users standing in front of the Kinect sensor and participate in the collaboration.

The relative positions of the users, the trophies on the table from the different views, and the fact that each user can interact with any of the trophies verify that the ASL is communicating the shared meshes and performing the synchronization correctly.

B. The CRCF Elements of the Collaborators

As discussed, the CRCF elements[10] for the Vive and the HoloLens configurations are: $\{O_v, P_v, S_{v+p}\}$ and $\{O_{v+p}, P_p, S_p\}$. The Kinect user interacts with virtual objects (O_v); has a point of view that is programmable (P_v) and as described, the Kinect user must perform actions in a physical space to interact with the virtual collaboration environment (S_{v+p}). In this way, the Vive CRCF signature is reproduced, and it becomes possible to investigate similar interaction parameters while trading-off immersion for a cost. As discussed previously, the CRCF does not differentiate between a VR and a *traditional* configuration, and thus the fourth collaborator has a CRCF signature of $\{O_v, P_v, S_v\}$.

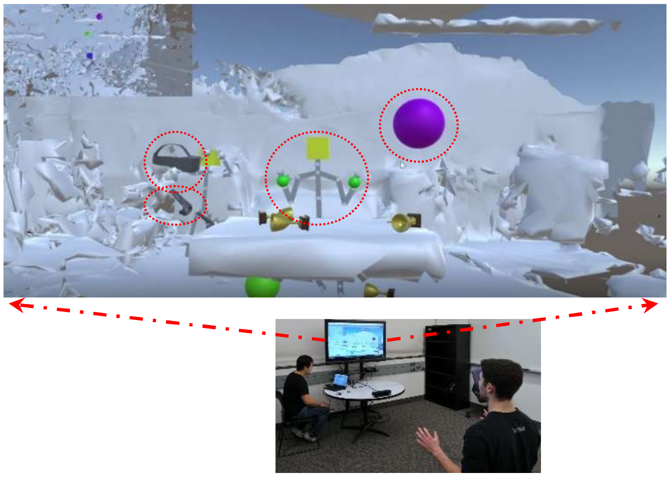


Fig. 4. View on the display of Kinect Collaborator

The CRCF provides a versatile platform for discussing, comparing, and contrasting the diverse technologies supporting the different collaborators. Individual CRCF elements provide useful insights into the potentials of a specific collaborator. For example, the S_{v+p} of the Kinect collaborator says that two persons can be presented in the same Kinect configuration and refer to their relative physical distance to critically evaluate the spatial situation of the virtual collaboration environment. Additionally, CRCF can be used as a platform for discussing alternative solutions. For example, in the absence of a costly AR device, one can ask if it is possible to construct a solution with a virtual point of view, or $\{O_{v+p}, P_v, S_p\}$. E.g., with the room mesh scanned and hidden, the environment in Figure 2 can be reproduced with a webcam running on a laptop, or viewing through a cell phone.

VI. CONCLUSION AND FUTURE WORK

Collaborations across distances and realities where participants are equipped with distinct technological configurations is an active and increasingly important area of study. The CRCF provides a platform for unifying the discussion and facilitates the understanding of these heterogeneous efforts where insights and opportunities can be derived. This paper presented results from an initial CRCF analysis demonstrating the potentials of the framework in supporting the understanding and differentiating technological configurations. More importantly, the insights based on the results of the analysis led to the investigation into the on-site remote presence sharing model, and the development of a Kinect application in an attempt to understand and reproduce the VR-like experience where immersion is traded-off for cost.

The continual investigations based on the results presented in Table 1 is important. At the elementary level, different

---

[10] The constant U_p element is omitted in the discussions.

permutations of virtual and physical elements should be examined and the implications to the corresponding CRCF signatures described. Other more purposeful directions of inquiry include, as illustrated in the case of the Kinect application, examining the potentials of reproducing VR/AR CRCF signatures by optimizing the degree of immersion; or examining the implications of replacing one of the elements of an established CRCF signature, e.g., replacing the physical point of view of an AR configuration with a virtual point of view. As pointed out earlier, this is one of the current undertakings at the authors' research group.

The complexity and functionality of VR/AR systems will continue to develop where resulting elements will continue to serve as the foundations for solutions to increasingly complex cross reality environments. In the current state, the CRCF can serve as a framework for a more designable, manageable, and teachable cross reality collaboration system. The CRCF will continue to evolve where it will serve as an essential platform for analyzing, comparing, contrasting and eventually predicting the performance and usability of cross reality collaboration systems.

ACKNOWLEDGMENT

The HoloLens device, which enabled our initial attempts, has been a generous loan from Jonathan Cults. Jason Pace ensured collaboration support stayed in focus. Avery Shinneman's ambitious to teach Geography with the Kinect helped started some of this work. Duncan MacMichael custom build our Vive machine.

REFERENCES

[1] Q. Zhao, "A survey on virtual reality," *Sci. China Ser. F Inf. Sci.*, vol. 52, no. 3, pp. 348–400, Mar. 2009.

[2] M. Billinghurst, A. Clark, and G. Lee, "A Survey of Augmented Reality," *Found. Trends® Human–Computer Interact.*, vol. 8, no. 2–3, pp. 73–272, 2015.

[3] J. Blascovich and J. Bailenson, *Infinite reality: avatars, eternal life, new worlds, and the dawn of the virtual revolution*, 1st ed. New York: William Morrow, 2011.

[4] "Augment Your Reality," *IEEE Comput. Graph. Appl.*, vol. 36, no. 1, pp. 6–7, Jan. 2016.

[5] C. Fleury, N. Férey, J. M. Vézien, and P. Bourdot, "Remote collaboration across heterogeneous large interactive spaces," in *2015 IEEE Second VR International Workshop on Collaborative Virtual Environments (3DCVE)*, 2015, pp. 9–10.

[6] H. Chen, A. S. Lee, M. Swift, and J. C. Tang, "3D Collaboration Method over HoloLens™ and Skype™ End Points," 2015, pp. 27–30.

[7] I. E. Sutherland, "A head-mounted three dimensional display," 1968, p. 757.

[8] J. H. Clark, "Designing surfaces in 3-D," *Commun. ACM*, vol. 19, no. 8, pp. 454–460, Aug. 1976.

[9] D. Harries, Ed., *The new media book*, Reprint. London: British Film Institute, 2004.

[10] M. Deering, "High resolution virtual reality," *ACM SIGGRAPH Comput. Graph.*, vol. 26, no. 2, pp. 195–202, Jul. 1992.

[11] R. Azuma, "A Survey of Augmented Reality," *Teleoperators Virtual Environ.*, vol. 6, no. 4, pp. 355–385, Aug. 1997.

[12] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Comput. Graph. Appl.*, vol. 21, no. 6, pp. 34–47, Dec. 2001.

[13] P. Hockett and T. Ingleby, "Augmented Reality with Hololens: Experiential Architectures Embedded in the Real World," *CoRR*, vol. abs/1610.04281, 2016.

[14] Feng Zhou, H. B.-L. Duh, and M. Billinghurst, "Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR," 2008, pp. 193–202.

[15] M. Sra and C. Schmandt, "Bringing real objects, spaces, actions, and interactions into social VR," in *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*, 2016, pp. 16–17.

[16] W. Liu, "Natural user interface- next mainstream product user interface," in *2010 IEEE 11th International Conference on Computer-Aided Industrial Design Conceptual Design 1*, 2010, vol. 1, pp. 203–205.

[17] O. Shaer and E. Hornecker, "Tangible user interfaces: past, present, and future directions," *Found. Trends Hum.-Comput. Interact.*, vol. 3, no. 1–2, pp. 1–137, 2010.

[18] R. Johansen, *Groupware: Computer support for business teams*. The Free Press, 1988.

[19] C. Blanchard *et al.*, "Reality built for two: a virtual reality tool," *ACM SIGGRAPH Comput. Graph.*, vol. 24, no. 2, pp. 35–36, Mar. 1990.

[20] J. Kaplan and N. Yankelovich, "Open Wonderland: An Extensible Virtual World Architecture," *IEEE Internet Comput.*, vol. 15, no. 5, pp. 38–45, Sep. 2011.

[21] Z. Lv, T. Yin, Y. Han, Y. Chen, and G. Chen, "WebVR——Web Virtual Reality Engine Based on P2P network," *J. Netw.*, vol. 6, no. 7, Jul. 2011.